Enhanced Neuro Fuzzy Approach For Detection of Credit Card Frauds

Manpreet kaur
Department of Computer Engineering & Technology
Guru Nanak Dev University
Amritsar,Punjab,India

Amit Chhabra
Department of Computer Engineering & Technology
Guru Nanak Dev University
Amritsar,Punjab,India

***Abstract*-The clustering mechanisms are devised to minimize the execution time associated with the sensors. To this category, chain based mechanisms such as stable election protocol is created. In this protocol entire network is divided into three parts. First part includes normal transaction. These transactions participate in detection of frauds if score attribute exceeds threshold value. Second transactions are known as least fraudulent if score attribute is within range. There are special node analyzer. This node monitor the execution time of conducted transaction. The proposed system integrated neural network based mechanism to eliminate execution time during the fraud detection phase. To accomplish the simulation is desired manner the rule base is defined. It has been analyzed about algorithm for without neural network based approach protocol became complex due to fuzzy technique involvement and thus further delay has been increased for fraud detection. To overcome this problem the algorithm can be divided in small problems using neural network techniques and the entire process can be enhanced by using NFBFD protocol in which neuro-fuzzy system involvement.**

Keywords*: *Fuzzy system, Social media, Clustering, NFBFD

I. INTRODUCTION

The application of [1]neural network is presented to determine frauds and its categorization in the proposed work. The dataset used in the proposed work is derived from the UCI website. The attributes of the dataset is not in compatible format. To bring this in compatible format, the attributes of dataset is adjusted according to requirements for accurately classifying the frauds. The neural network based mechanism is used using the MATLAB software. The entire mechanism of fraud detection is partitioned into phases. The phases of the proposed work is partitioned into sections. These sections includes pre-processing, semantic based clustering , segmentation and then classification.

The[2] pre-processing phase is used to eliminate noise if any from within the dataset. The dataset may contain noise or irrelevant data. The noise may be introduced due to elimination of extra digits from the numeral data. [3]In order to eliminate the noise if any from the data, normalization procedure and insignificant value removal procedure is incorporated.

After pre-processing mechanism[4] clustering procedure by eliminating stop words and key word sensitive mechanism is employed. The keyword sensitive mechanism group together words or numerals having similar context within same group. This is done in order to reduce the execution time required to detect the fraud if any within the transactions.

[5]Segmentation rejects the critical regions from the non critical regions or groups. This mechanism allows only necessary data to be focused upon through the proposed work. This mechanism is implemented using neural network based approach.

Classification is performed by using rules of neural based approach. This approach determines the distinct categories of frauds. The primary parameter for the detection of fraud is score. In addition ‘number of clicks’ indicating amount of time transaction is attempted to take place again and again is critical in severity of attack. In case score is lower than threshold value then fraud is not detected and transaction is labeled as clean.

Rest of the paper is organized as under: section 2 gives the literature survey involving mining and neural network based approach, section 3 gives the problem definition and objective of study, section 4 gives the methodology of work, section 5 gives the performance analysis and result, section 6 gives conclusion.

II. LITERATURE SURVEY

This section describes the mechanism used to determine the frauds within the existing system. The techniques include mixture of data mining and neural network based approaches. The survey includes transaction safety, credit card frauds etc.

In the thorough review of various data mining based techniques [6]proposed a mechanism to analyze big data corresponding to transaction safety monitoring. Architectural framework for transaction safety monitoring using big data technology with high precision tackled through this literature.
[7]that are used for analysis the data of shifting degree of transaction and then use various data mining techniques for evaluating this data. In this literature various clustering algorithms are also overviewed and these algorithms are based on cluster processing that inspects usage decisions. It is mainly based on two fundamental decisions that use separation metric in time span and this time arrangement is used for characterization.
The mean time based approach [8] is used that are used to develop prediction based on the fluctuation that are gathered from transaction. Various researches used clustering based approaches to formulate decisions about transaction safety. It uses separation metric that are created in particular time span and then gives “medoid” on the basis of agent focus. In this time based correlation and data mining techniques are used that render comparable arrangement for separating data.
The DTW approach [9] conveys data about the closeness and move between two time arrangements, thus represents something like one huge reason for blunder in time arrangement data mining. In a review of time-arrangement clustering strategies, it detailed that the vast majority of the distinguished systems can be gathered by their utilization of the first data. After utilizing clustering and data mining system it gives various abnormal state that are specified in this dataset. The clustering mechanism is used for locating neighborhood and it does not use any adjustment strategy. It also extract data from datasets and after that clustering is applied that separate the highlighted data from this dataset.
It [10] proposed the model that uses parameters of transaction data and then patterns, practices are used. Then again, the age of various models in subsets or cases of the data take into account the utilization of clustering on model parameters.
[11]Lee et al. 2015 proposed a mechanism to process data stream with the use of distributed computing platform. First data stream channel is push and pull transfer protocol and second is pub and sub transfer protocol. In this literature push is used to distribute packets among all pulls. Pull act as data receiving node. For fast path stream channel multicast protocol is designed in this literature. Overall effective protocol reduce execution speed for data stream processing.
[12]McHugh et al. 2018 for big data processing, knowledge based framework is proposed. Data could be images, time series, unstructured text etc. To handle any kind of data that becomes need of the hour and used now days, Semantic toolkit is used in this literature. In this literature data capturing mechanism depends upon graph structure and hence interactive data processing mechanism was proposed.
[13] Taleb and Serhani 2017 To eliminate anomalies from the data and make it feasible to take decision through the presented data, big data preprocessing mechanism is proposed It gives quality based rule model that will set quality requirement and apply big data for evaluating quality. In this literature the rules are generated and based on these rules, quality is tested. The results show that quality is improved. This literature is capable of validating and checking quality of data with optimality.
[14] Barber et al. 2017 in this the proposed system discussed the role of database in the analysis of big data. Hybrid transactional and analytical wildfire system presented for analysis through considered approach. Complex analytical requests are handled by Spark eco system. Queries which are handled through this literature include insert, update and delete. Through this system, Availability causes degradation in performance This anomaly becomes overwritten using replication mechanism.
[15] Poledna et al. 2015 for risk assessment in big data analysis, agent based approach is proposed. Basic and advanced requirements required establishing and use big data analytics becomes critical in this discussion.

Systematic risk assessment through the use of parallel and supercomputing presented in graphical form through this literature.

[16] Jung et al. 2017 accountable protocol is used to detect frauds in trading. The proposed protocol ensure fair trading and performed automatic accountability check. To evaluates the trade sold by seller to check for re-sell trades, Uniqueness index method employed in this protocol. Proposed model verified through the use of ProVerif and rigorous analysis mechanism.

[17] Mccormack and Smyth 2017 for string processing in big data, mathematical solution is proposed in this literature. Identity correlation mechanism primarily used for matching contents against big datasets. Mathematical tool are used in Identity correlation mechanism to match the given word with entire dataset. By the use of this approach, the complexity of searching is reduced.

The literature survey is concluded by stating that least amount of work is done using neural network based approach to detect fraud if any from within the transactions.

III. PROBLEM DEFINITION AND OBJECTIVES

Existing approach of bot detection suffer from multiple problems, all these problems are listed as follows:

- Potentially huge set of candidate sequences may cause problem and data may goes unidentified.
- Multiple scans of databases causes huge execution time.
- Difficulties at mining long sequential patterns and hence classification accuracy is a problem.

Mechanism to be used to solve above listed problems

Modifed ANN algorithm for discovering abnormal pattern detection.

In Prefix span patterns is discovered by arranging the terms in ascending oder of occurance. In proposed mechanism analysis can be conducted by using mechanism to check data both from ascending order as well as descending order of sequence.

The objecve of the proposed work is given as under

- To overcome the problem of latency in detecting attacks.
- To detect attack with high classification accuracy.
- To detect multiple attack with high quantity within prescribed time period.
- To compare existing bot based approach and proposed ANN approach in attack detection.

Next section gives the propsoed work with methodology and rule based used for the detection of frauds.

IV. METHODOLOGY OF WORK

The methodology of the proposed work consist of Pre-processing, clustering , segmentation and classification mechanism. The pre-processing mechanism includes handling noise from within the dataset. The noise could be unnormalised dataset. The normalization mechanism eliminate the undesired data from the dataset derived from the UCI website. The mechanism employed for clustering includes semantic based keyword procedure, segmentation is on the basis of ANN and classification uses the rule based of fuzzy.

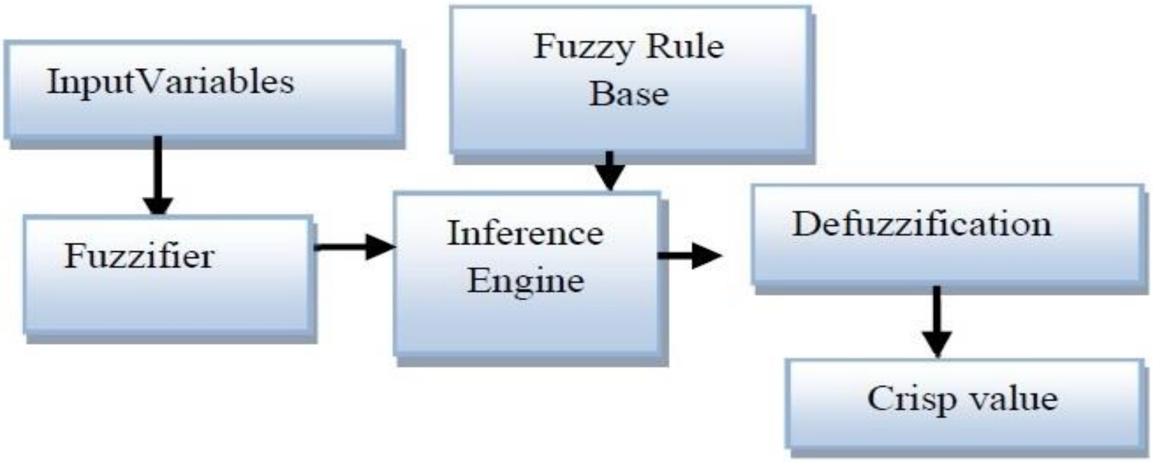


Figure 1: Fuzzy based engine for fraud detection

The fuzzy based engine is designed using fuzzy designer engine of MATLAB. The four parameters used for this purpose includes centrality or number of previous attempts, distance or number of clicks, score or energy and concentration that is labelled as assessment in the proposed work.

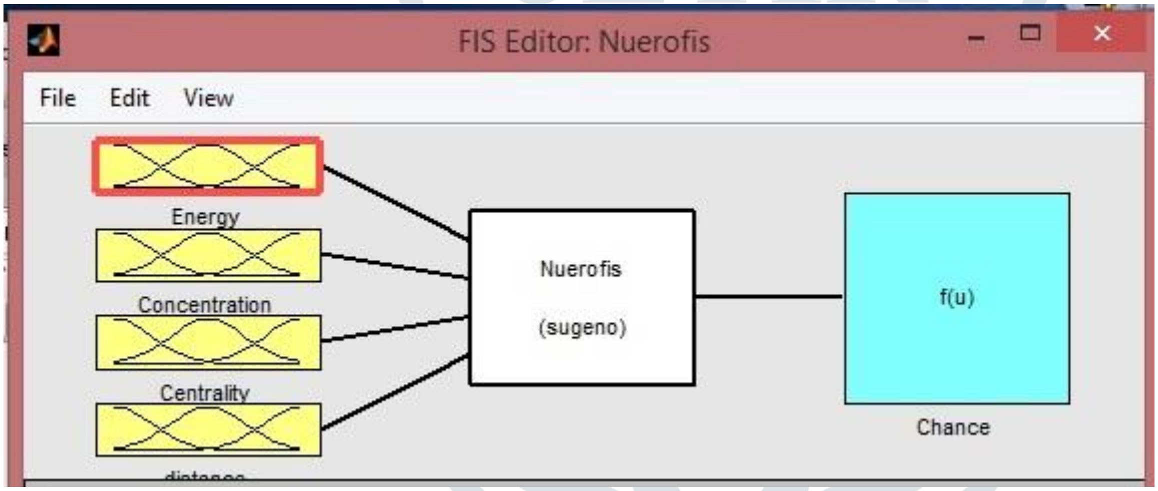


Figure 2: Input and output variables of the propsoed work

The rule base for the proposed system is categorised into 32 rules. In the proposed NFLAD system, 32 rules are used in fuzzy inference (Table 1)

Chance_of_fraud = (Energy-1) + Concentration + Centrality + Distance

From equation , consider energy is equal to energy-1 because in every round there is some energy consumption. Chance value is calculated by the sum of all the linguistic variables. Output of the chance is further divide into seven parts as Very small (vS), small (s), rare small (rS), medium (M), rare large (rL), large (L), very large (vL) with different membership functions.

RULE BASE

The rule based for proposed attack detection is created within the neurofuzzydesigner using MATLAB. The interface provide graphical user interface for the construction of the rule base to be used with the proposed system. For score and assesment: L(low), M(medium), H(high)is used. In the number of previous attempts: C(close) i.e value close to the threshold value, A(average), F(far). Number of clicks can be represented by: N(not done), M(in middle), F(full). The rule based is given as under:

S No	Score	Assessment	Number of previous attempts	Number of Clicks	Chance
1	L	L	C	N	S
2	L	L	C	M	S
3	L	L	C	F	vS
4	L	L	A	N	S
5	L	L	A	M	S
6	L	L	A	F	vS
7	L	L	F	N	M
8	L	L	F	M	M
9	L	L	F	F	L
10	L	M	C	N	S
11	L	M	C	M	S
12	L	M	C	F	vS
13	L	M	A	N	S
14	L	M	A	M	S
15	L	M	A	F	vS
16	L	M	F	N	M
17	L	M	F	M	M
18	L	M	F	F	L
19	L	H	C	N	rS
20	L	H	C	M	S
21	L	H	C	F	vS
22	L	H	A	N	rS
23	L	H	A	M	rS
24	L	H	A	F	S
25	L	H	F	N	rL
26	L	H	F	M	M
27	L	H	F	F	L
28	M	L	C	N	M
29	M	L	C	M	M
30	M	L	C	F	rS
31	M	L	A	N	rS
32	M	L	A	M	rS

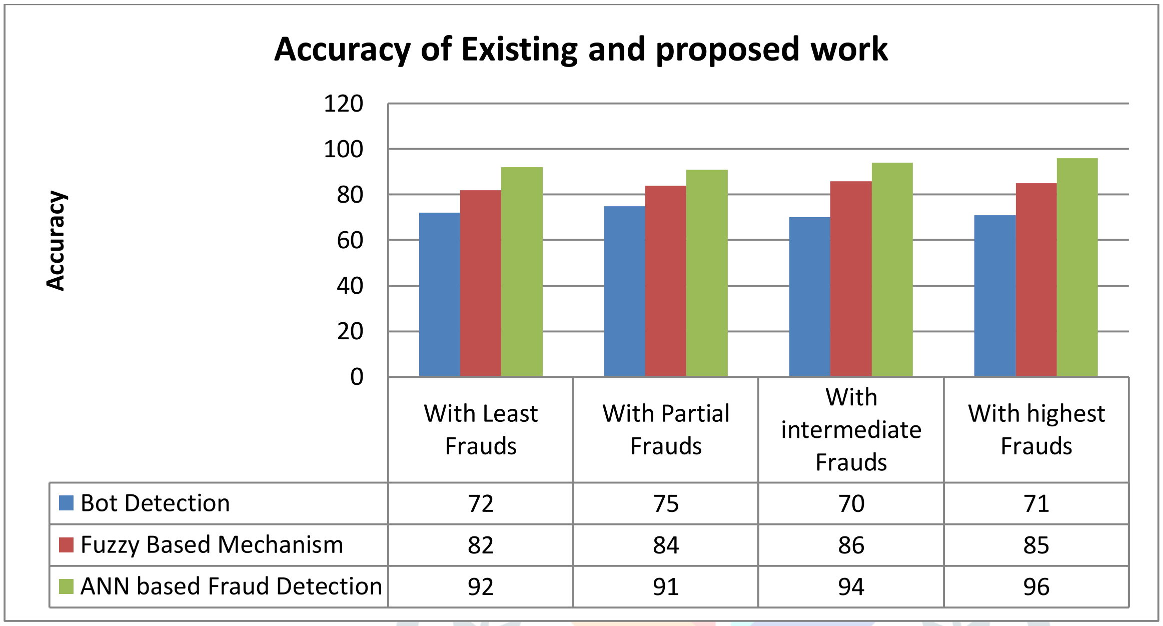
Table 1: Rule base for NFL based fraud detection

The mechanism used provides best possible result in terms of classification accuracy and number of fraud patterns discovered. Next section gives the perforamnce evaluation of the existing and proposed system.

V. PERFORMANCE EVALUATION AND RESULTS

The result obtained through the proposed system is better in terms of execution time and number of frauds discovered. The mechanism obtain four classes of results. The first class does not detect any frauds, second class detect partial frauds, third class specifies medium frauds and forth class detect high frauds.

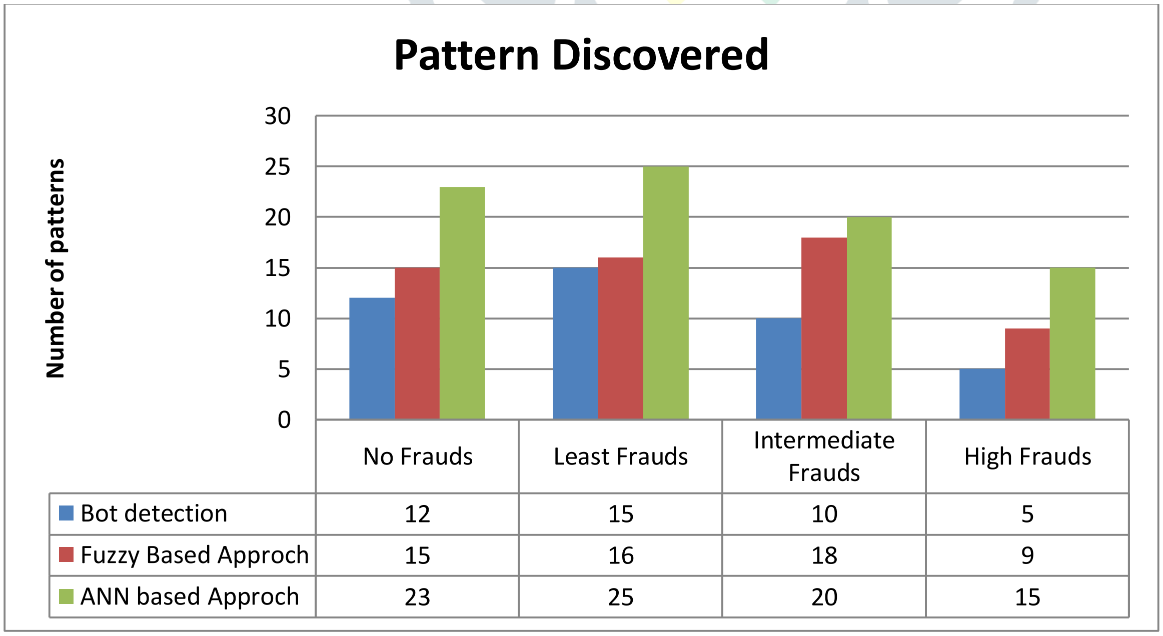
The figure 3 indicates the frauds detected without and with the use of ANN mechanism.



	With Least Frauds	With Partial Frauds	With intermediate Frauds	With highest Frauds
🟦 Bot Detection	72	75	70	71
🟥 Fuzzy Based Mechanism	82	84	86	85
🟩 ANN based Fraud Detection	92	91	94	96

Figure 3: Classification Accuracy

Figure 3 indicates the betterment of proposed work by significant margin. The margin of accuracy is improved by 8%. In addition result in terms of fraud patterns discovered is also improved. The pattern discovered through the proposed approch is given through figure 4.



	No Frauds	Least Frauds	Intermediate Frauds	High Frauds
🟦 Bot detection	12	15	10	5
🟥 Fuzzy Based Approch	15	16	18	9
🟩 ANN based Approch	23	25	20	15

Figure 4: Number of frauds detected

The proposed mechanism gives the better result in terms of number of frauds discovered. The mechanism also improves execution time. The number of patterns discovered are much higher as compared to existing approach.

VI. CONCLUSION

The proposed approch using the application of ANN gives better result in terms of classification accuracy. The classification accuracy is acheved by determine the class of transaction. There are four classes in which given dataset is partitoned. The four classes divides the entire dataset into segments and if some tuple cannot be categorised into one of the class than error appears that decreases classification accuracy. In addition number of frauds discovered is much higher as compared to fuzzy and bot detection mechanism. The result improvement by 8% is observed which is significant and proves the worth of study.

VII. REFERENCES

[1]T. G. T. Guo and G.-Y. L. G.-Y. Li, “Neural data mining for credit card fraud detection,” *2008 Int. Conf. Mach. Learn. Cybern.*, vol. 7, pp. 1–4, 2008.

[2]N. Upasani and H. Om, “Evolving fuzzy min-max neural network for outlier detection,” *Procedia Comput. Sci.*, vol. 45, no. C, pp. 753–761, 2015.

[3]L. Park, “Learning of Neural Networks for Fraud Detection Based on a Partial Area Under Curve,” *Int. Symp. Neural Networks*, vol. 3497, pp. 922–927, 2017.

[4]A. Nazemi and A. Maleki, “Artificial neural network classifier in comparison with LDA and LS-SVM classifiers to recognize 52 hand postures and movements,” *Proc. 4th Int. Conf. Comput. Knowl. Eng. ICCKE 2014*, pp. 18–22, 2014.

[5]S. M. Zhou, M. A. Rahman, M. Atkinson, and S. Brophy, “Mining textual data from primary healthcare records: Automatic identification of patient phenotype cohorts,” *Proc. Int. Jt. Conf. Neural Networks*, pp. 3621–3627, 2014.

[6]B. Li, X. Ming, and G. Li, “Big Data Analytics Platform for Flight Safety Monitoring,” pp. 350–353, 2017.

[7]G. Zhu, K. Song, and P. Zhang, “A Travel Time Prediction Method for Urban Road Traffic Sensors Data,” *2015 Int. Conf. Identification, Information, Knowl. Internet Things*, pp. 29–32, 2015.

[8] S. Jasra, J. Gauci, A. Muscat, and G. Valentino, “Literature review of machine learning techniques to analyse flight data,” *Res. Gate*, no. October, 2018.

[9] G. Li, T. Yuan, S. J. Qin, and T. Chai, “Dynamic time warping based causality analysis for root-cause diagnosis of nonstationary fault processes,” *Int. J. Autom. Control*, pp. 1289–1294, 2015.

[10] V. M. Janakiraman and D. Nielsen, “Anomaly Detection in Aviation Data using Extreme Learning Machines,” 2016.

[11] Y.-J. Lee, M. Lee, M.-Y. Lee, S. J. Hur, and O. Min, “Design of a scalable data stream channel for big data processing,” *2015 17th Int. Conf. Adv. Commun. Technol.*, pp. 537–540, 2015.

[12] J. McHugh, P. E. Cuddihy, J. W. Williams, K. S. Aggour, V. S. Kumar, and V. Mulwad, “Integrated access to big data polystores through a knowledge-driven framework,” *Proc. - 2017 IEEE Int. Conf. Big Data, Big Data 2017*, vol. 2018-Janua, pp. 1494–1503, 2018.

[13] I. Taleb and M. A. Serhani, “Big Data Pre-Processing: Closing the Data Quality Enforcement Loop,” *Proc. - 2017 IEEE 6th Int. Congr. Big Data, BigData Congr. 2017*, no. 1, pp. 498–501, 2017.

[14] R. Barber, C. Garcia-arellano, R. Mueller, A. Storm, G. Lohman, C. Mohan, and H. Pirahesh, “Evolving

Databases for New-Gen Big Data Applications," *Cidr*, no. 3, 2017.

[15] S. Poledna, M. G. Miess, S. Schmelzer, E. Rovenskaya, S. Hochrainer-stigler, and S. Thurner, "Agent-based Modelling of Systemic Risk : A Big-data Approach Application : Systemic Risk Triggered by Natural Disasters Big-data," *Sebastian Poled. Michael Greg. Miess, Stefan Schmelzer*, p. 10, 2015.

[16] T. Jung, X. Y. Li, W. Huang, J. Qian, L. Chen, J. Han, J. Hou, and C. Su, "AccountTrade: Accountable protocols for big data trading against dishonest consumers," *Proc. - IEEE INFOCOM*, 2017.

[17] K. Mccormack and M. Smyth, "A Mathematical Solution to String Matching for Big Data Linking," *J. Stat. Sci. Appl.*, vol. 5, pp. 39–55, 2017.